Powered by

Futurense

Module 02

Types of Correlation

- Correlation is a measure of how two variables change together.
- A correlation coefficient quantifies this relationship’s strength and direction.

Learning Outcomes

List five different types of correlation coefficients used in statistics

Identify the correct usage of these correlation coefficients based on the data type

Compute the appropriate correlation between variables

Topics

Pearson's Correlation Coefficient

Spearman Rank Order Correlation

Kendall's Tau

Point Biserial Correlation

Phi coefficient

Topics

Pearson's Correlation Coefficient

Spearman Rank Order Correlation

Kendall's Tau

Point Biserial Correlation

Phi coefficient

Pearson's correlation measures the linear relationship between **two continuous variables.**

It assumes that the data is normally distributed and that the relationship is linear.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \cdot \sqrt{\sum(Y_i - \bar{Y})^2}}$$

E.g. Study hours—Exam Scores
Exercise hour—Heart rates

Spearman Rank Order Correlation

Spearman's correlation assesses the monotonic relationship between two variables **based on rank orders** rather than raw values.

It is useful when data is ordinal or when the relationship is not strictly linear.

Does not assume normality; less affected by outliers.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Where d_i is the difference between ranks.

Spearman Rank Order Correlation

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where d_i is the difference between ranks.

Group Study Hours vs. Presentation Quality

Spearman Rank Order Correlation

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where d_i is the difference between ranks.

$$\rho = 1 - \frac{6 \Sigma \left(1^2 + (-1)^2\right)}{5(5^2 - 1)}$$

$$= \underline{\underline{0.9}}$$

Student	Group Study Hours	Ranks	Pres. Quality Rank (Ordinal)	d_i
A	4	2	2	0
B	6	4	3	1
C	2	1	1	0
D	7	5	5	0
E	5	3	4	-1

Strong monotonic trend — students who study more in groups tend to perform better in presentations.

Kendall’s Tau

Measures **ordinal association** between **two ranked variables.**

Counts **concordant** and **discordant** pairs.

Preferred for small datasets or with many tied ranks.

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}$$

Where:

- **C** = number of concordant pairs
- **D** = number of discordant pairs

Kendall's Tau

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}$$

Where:

- **C** = number of concordant pairs
- **D** = number of discordant pairs

ID	Rank in Math (X)	Rank in Science (Y)
A	1	2
B	2	3
C	3	1
D	4	4

A pair of observations (i, j) is **concordant** if:

The ranks of both elements **move in the same direction.**

- If $X_i > X_j$ and $Y_i > Y_j$, or
- If $X_i < X_j$ and $Y_i < Y_j$

Example:
Student P scores higher in **both math and science** than Student Q → **Concordant**

Kendall's Tau

ID	Rank in Math (X)	Rank in Science (Y)
A	1	2
B	2	3
C	3	1
D	4	4

Pair	(X_i, X_j) vs (Y_i, Y_j)	Comparison Result
(A, B)	(1, 2) and (2, 3)	Concordant
(A, C)	(1, 3) and (2, 1)	Discordant
(A, D)	(1, 4) and (2, 4)	Concordant
(B, C)	(2, 3) and (3, 1)	Discordant
(B, D)	(2, 4) and (3, 4)	Concordant
(C, D)	(3, 4) and (1, 4)	Concordant

$$\boxed{\begin{aligned} C &= 4 \\ D &= 2 \end{aligned}}$$

The total number of unique pairs is

$$\frac{n(n-1)}{2} = \frac{4 \cdot 3}{2} = 6$$

✓Kendall's Tau

ID	Rank in Math (X)	Rank in Science (Y)
A	1	2
B	2	3
C	3	1
D	4	4

Pair	(X_i, X_j) vs (Y_i, Y_j)	Comparison Result
(A, B)	(1, 2) and (2, 3)	Concordant
(A, C)	(1, 3) and (2, 1)	Discordant
(A, D)	(1, 4) and (2, 4)	Concordant
(B, C)	(2, 3) and (3, 1)	Discordant
(B, D)	(2, 4) and (3, 4)	Concordant
(C, D)	(3, 4) and (1, 4)	Concordant

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}$$

Where:

- **C** = number of concordant pairs
- **D** = number of discordant pairs

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)} = \frac{4-2}{6} = \frac{2}{6} = 0.33$$

Moderate positive association

✓ Point-Biserial Correlation

Measures correlation between a **binary** and a **continuous** variable.

Special case of Pearson's *r*.

$$r_{pb} = \frac{\bar{X}_1 - \bar{X}_0}{s} \cdot \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Where:

- $\bar{X}_1, \bar{X}_0$: Means of groups 1 and 0
- Sample s: Standard deviation of all scores

$\boxed{r_{pb} \approx 0.98}$ ← close to 1

Attendance in revision session vs. Exam Score

Student	Attended (1/0)	Exam Score
A	1	88
B	1	85
C	0	70
D	0	72
E	1	90

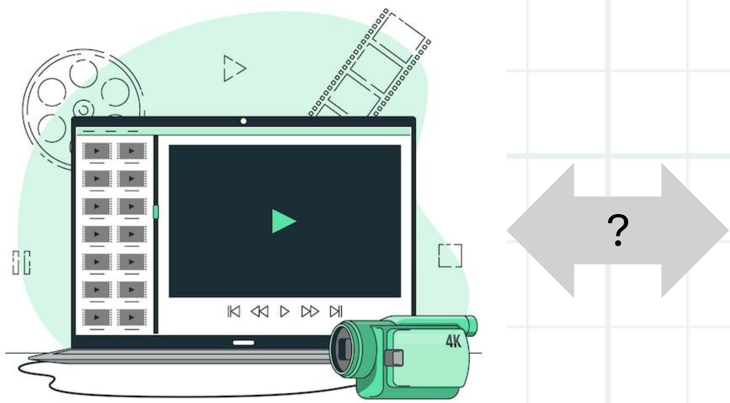
Attending revision sessions is highly associated with better scores.

Phi Coefficient

Measures association between **two binary variables**.

Equivalent to Pearson's *r* on a 2 x 2 table

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$



Video lectures watched
(Yes/No)

Exam
(Pass/Fail)

Phi Coefficient

Measures association between **two binary variables**.

Equivalent to Pearson's *r* on a 2 x 2 table

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

	Passed	Failed
Watched Video	8 = a	2 = b
Didn't Watch	3 = c	7 = d

$$a = 8, b = 2, c = 3, d = 7$$

$$\phi = \frac{(8\times7-2\times3)}{\sqrt{(10)(10)(11)(5)}} \approx 0.53$$

Moderate association — watching tutorial videos improves the likelihood of passing.

Summary

Data Type	Use Case	Correlation	Key element in the correlation
Continuous Continuous	Linear relationship	**Pearson**	Covariance and standard deviation
Ordinal or Ranked	Monotonic relationship	**Spearman**	Rank differences
	Rank concordance	**Kendall's Tau**	Concordant vs. Discordant
Binary Continuous		**Point-Biserial**	Group mean difference / standard deviation
Binary Binary		**Phi coefficient**	2×2 table counts

Online Material

Correlation analysis, *DataTab*

https://youtu.be/G5FkaxWBtkM